Age of the Common Ancestor of Human Mitochondrial DNA

Masatoshi Nei
Institute of Molecular Evolutionary Genetics and Department of Biology, Pennsylvania State University

Studying the evolutionary change of the control region of human (H) and chimpanzee (C) mitochondrial DNAs (mtDNAs), Vigilant et al. (1991) concluded that the root of the phylogenetic tree for H mtDNAs is located at a sequence from African populations and that the age of the common H ancestral mtDNA is 166,000–249,000 years. Hedges et al. (1992) and Templeton (1992) challenged the first of these conclusions, showing that valid statistical tests do not support it. In this letter, I would like to show that Vigilant et al.'s second conclusion is also questionable and that, if we consider statistical errors associated with the rate of nucleotide substitution, the range of the estimate of the age becomes much wider.

In the control region, transitional nucleotide substitution is known to occur at a much higher rate than does transversional substitution, and the transitional differences between the H and C mtDNAs seem to have reached a saturation level at hypervariable sites of this region (Kocher and Wilson 1991). Vigilant et al. therefore used information on the transversional differences between H and C and the transition/transversion ratio obtained from H data, to calibrate the rate of nucleotide substitution.

Let \( n \) be the number of transversional differences between H and C, and let \( R \) be the ratio of transitional (\( s \)) and transversional (\( v \)) substitutions observed among H sequences. (Hasegawa and Horai 1991) suggested that \( R \) may be different between H and C, but K. Tamura's (personal communication) detailed analysis indicates that it is essentially the same for the two species. Here I consider only H data, following Vigilant et al.] The expected total number of nucleotide substitutions between H and C can then be estimated by \( n + nR \). So, if the total number of nucleotides examined is \( m \), then the total number of nucleotide substitutions, per site, between H and C is given by

\[
d = \frac{n + nR}{m} = p(1 + R),
\]

where \( p \) is \( n/m \) and is assumed to be small (say, \(<0.10\)). In Vigilant et al.'s case, \( n = 26.4, R = 15 \), and \( m = 610 \). Therefore, they obtained \( d = 0.692 \). They assumed that the divergence time between H and C is 4–6 Myr. Thus, the rate of nucleotide substitution per site per two lineages (\( r = 2\lambda \)) was 11.5%–17.3% per Myr. If we use the standard measure of substitution rate, i.e., the rate per site per lineage (\( \lambda \)), then the rate becomes \( \lambda = 5.75 \times 10^{-8}–8.65 \times 10^{-8} \) per year per lineage.

Let us now compute the standard error of the total number \( d \) of nucleotide substitutions. If we use the delta technique in statistics (Stewart and Ord 1987, p. 324), the variance of \( d \) is given by

\[
V(d) = p^2V(R) + (1+R)^2V(p) + 2p(1+R)Cov(p,R),
\]

where \( V(R) \) and \( V(p) \) are the variances of \( R \) and \( p \), and \( Cov(p,R) \) is their covariance. The covariance is estimated as

\[
\text{Cov}(p,R) = \frac{\text{Var}(p) \cdot \text{Var}(R)}{m},
\]

where \( \text{Var}(p) \) and \( \text{Var}(R) \) are the variances of \( p \) and \( R \), respectively.


1. Key words: molecular clock, mitochondrial D-loop, human divergence, confidence interval.

Address for correspondence and reprints: Masatoshi Nei, Institute of Molecular Evolutionary Genetics and Department of Biology, Mueller Laboratory, Pennsylvania State University, University Park, Pennsylvania 16802.

© 1992 by The University of Chicago. All rights reserved.
0737-4038/92/0906-0012$02.00
where $V(x)$ and $\text{Cov}(x,y)$ denote the variance of $x$ and the covariance of $x$ and $y$, respectively. In the present case, $p$ and $R$ are independent of each other, so that $\text{Cov}(p,R) = 0$. Furthermore, we have $V(p) = p(1-p)/m$, and

$$V(R) = V(s/v) = \frac{s^2}{v^2} \left[ \frac{V(s)}{s^2} + \frac{V(v)}{v^2} - \frac{2\text{Cov}(s,v)}{s \cdot v} \right],$$

(3)

where $s$ and $v$ are the numbers of the transitions and transversions observed in $H$ data, respectively.

Since $s$ and $v$ are obtained, by a parsimony method, from comparison of closely related $H$ sequences, they are approximately Poisson distributed, and thus $V(s) = s$ and $V(v) = v$. Theoretically, $s$ and $v$ are expected to be slightly negatively correlated, but, for closely related sequences, $\text{Cov}(s,v)$ is virtually 0. Vigilant et al. state that they obtained $s+v = 528$ and $s/v = 15$. Therefore, $s$ and $v$ must be 495 and 33, respectively. We then have $V(R) = 7.27$. We also have $p = 0.0433$ and $V(p) = 6.7878 \times 10^{-5}$. Thus, $V(d) = 0.0310$, and the standard error $(s_d)$ of $d$ is 0.176.

This indicates that the 95% confidence interval of $d$ is $d \pm 2s_d = 0.340 - 1.044$ and that, if the divergence time between $H$ and $C$ is 4–6 Myr, then the rate of nucleotide substitution ($r$) is 5.7%–26.1% per Myr. Vigilant et al. state that the ancestor corresponding to the deepest node of the phylogenetic tree for $H$ sequences is placed at $d_c = 2.87\%$ on the scale of accumulated sequence differences. Therefore, if we use the above range of the $r$ value, the age of the common ancestral mtDNA is estimated to be 110,000–504,000 years. [Theoretically, $d_c$ is subject to statistical errors, but this factor has not been considered here because $d_c$ is highly correlated with $s$ and $v$.]

If we consider this factor properly, the range of the estimate of the age of the common ancestor would slightly increase. Note also that the above estimate of the age has been derived on a statistical basis and is different from a mere conjecture presented by Stoneking and Cann (1989).]

Analyzing sequence data obtained by Vigilant et al. (1989), Horai and Hayasaka (1990), and others, Hasegawa and Horai (1991) estimated that the 95% confidence interval of the age of the common ancestor is 180,000–380,000 years. This estimate was obtained from a comparison of the $H$ and $C$ sequences by using a specific mathematical model. Our recent study (K. Tamura and M. Nei, unpublished data) has shown that this model is unrealistic because it does not take into account the continuous variation of substitution rate among sites of the control region (Kocher and Wilson 1991). Therefore, Hasegawa and Horai's estimate seems to be less reliable.

At present, there is a heated controversy over the origin of *Homo sapiens*. Cann et al. (1987), Stringer and Andrews (1988), and others maintain the view that *H. sapiens* originated in Africa ~200,000 years ago and later spread through the world (hypothesis of an African origin). By contrast, Wolpoff et al. (1984), Wolpoff (1989), and others believe that *H. sapiens* evolved gradually from *H. erectus* in various places of the world during the last 1 Myr, with the help of gene migration between different populations (hypothesis of multiregional evolution). Vigilant et al. (1991) took their data as support of the hypothesis of an African origin. As mentioned earlier, Hedges et al.'s (1992) and Templeton's (1992) reexamination of their phylogenetic analysis does not necessarily support the African-origin hypothesis. The present calculation shows that their estimate of the age of the ancestral mtDNA is also unreliable and that the discrimination between the two hypotheses is not as clear-cut as they thought. Nevertheless, it is important to note that the mtDNA data still suggest a relatively recent origin in Africa [see the neighbor-joining tree constructed by Hedges et al.].
To settle these problems, it is necessary to investigate the evolutionary history of polymorphic alleles at many independently evolving genetic loci (Nei and Livshits 1989, 1990; Bowcock et al. 1991).

Acknowledgments

I thank Koichiro Tamura and Mark Stoneking for helpful discussions. This study was supported by research grants from the National Institute of Health and the National Science Foundation.

LITERATURE CITED


Michael Bulmer, reviewing editor

Received March 30, 1992; revision received May 22, 1992

Accepted May 22, 1992